

学校编码: 10384

分类号_____密级_____

学号: X200417004

UDC _____

厦 门 大 学

硕 士 学 位 论 文

现代教育测量理论在标准参照语言测试中的应用与案例研究

(A Study and Case Analysis of the Application of Modern Educational Measurement Theory in Criterion-referenced Language Testing)

黄 锐

指导教师姓名: 史 秋 衡 教授

专 业 名 称: 高 等 教 育 学

论文提交日期: 2007 年 10 月

论文答辩日期: 2007 年 月

学位授予日期: 2007 年 月

答辩委员会主席: _____

评 阅 人: _____

2007 年 10 月

厦门大学博硕士论文摘要库

厦门大学学位论文原创性声明

兹呈交的学位论文，是本人在导师指导下独立完成的研究成果。本人在论文写作中参考的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文产生的权利和责任。

声明人（签名）：

年 月 日

厦门大学博硕士论文摘要库

厦门大学学位论文著作权使用声明

本人完全了解厦门大学有关保留、使用学位论文的规定。厦门大学有权保留并向国家主管部门或其指定机构送交论文的纸质版和电子版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆被查阅，有权将学位论文的内容编入有关数据库进行检索，有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

本学位论文属于

1、保密（ ），在 年解密后适用本授权书。

2、不保密（ ）

（请在以上相应括号内打“√”）

作者签名：

日期： 年 月 日

导师签名：

日期： 年 月 日

厦门大学博硕士论文摘要库

内容提要

高等教育质量保障是高等教育理论研究中的关键问题,要保证多元化高等教育质量,考试作为衡量教育质量的主要手段,其方式、命题和评分在教学中起着至关重要的作用,科学化的考试能有效推动学生发展,规范和引导教师的教学行为,达到培养规格及教学计划的要求。本论文以检测教学是否按教学大纲为目标的标准参照性语言测试为研究对象,通过阐述基于教育测量理论的标准参照性测验的原理,与常模参照性测验相比,说明高校教学更应注重以检测学生“掌握与未掌握”为目标的标准参照测验,并通过对集美大学英语专业四级考试结果的具体分析,实证说明用 A 、 B 、 Φ 值和项目反应理论的三个参数,更科学地检测学生对知识的掌握程度,得出为确保标准参照语言测试对外语教学的正反拨作用,提高命题的质量,保障高等教育教学质量非常重要这一结论。

论文共分为四部分。

第一章绪论主要介绍了论文的选题、概念的界定、研究现状和研究思路。

第二章详尽地阐述了基于传统教育测量理论下的标准参照测验基本理论。通过解释标准参照测验的兴起与内涵、标准参照语言测试的编制原则与使用,并与常模参照测验对比,说明两种测验的异同,同时提出另一种描述分数解释方法,以及标准参照语言测试的高校教学质量测量法则,作为后文的理论基础。

第三章通过对试题分数各统计值的描述和介绍,区分了常模参照和标准参照测验的各统计值的含义,进而介绍影响标准参照测验的 A 、 B 和 Φ 值对试题设计的指导,以及代表现代教育测量理论的项目反应理论的优势对标准参照语言测试的作用,最后用标准参照语言测试案例来实证项目反应理论的三个参数对测验的影响。

第四章用实际英语专业四级考试案例进一步说明项目反应理论的三个参数对标准参照语言测试试题的影响,总结出相比标准参照测验的 A 、 B 和 Φ 值,三个参数的数值更科学,对试题的筛选和修正更精确,从而在提高命题质量,甄别不同能力的学生,保障高等教育教学质量提供了科学依据。

结语对研究的新意和主要结论及有待继续研究的问题作了一个简单的交代。

关键词: 现代教育测量理论; 标准参照语言测试; 英语专业四级考试; 应用

厦门大学博士论文摘要库

Abstract

Higher Education Quality Assurance is a key issue in the theory of higher education. This includes the multi-evaluation of the quality of higher education, paper design, method of testing, score description-analysis, and the essential role of examination in teaching as a main assessment-evaluation for quality education.

This study which used the Criterion-referenced Language Testing (CRLT) method aimed to examine a syllabus based language teaching. It explained the principles of CRLT and its comparison to the principles of Norm-Referenced Test (NRT) to explicate the similarities and differences between the two tests; a distinction on NRT and CRLT scores that described and presented the statistical data; an analysis on the result of Test English as a Major Grade 4 (TEM4) taken by students in Jimei University to improve CRLT; a CRLT assessment principle of the quality of Higher Education; the use of A, B, Φ values, Item Response Theory (IRT) to guide CRLT items test design and to protect the quality assurance of Higher education.

The dissertation comprises four parts.

Chapter one gave a brief introduction of the main reasons for choosing this topic, literature review, method of research and identification of several concept definitions.

Chapter two described in details the principles of CRLT based on the traditional theory of educational measurement. This part also described the interpretation and comparison of the contents, the basic ideas and methods of CRTs and NRTs to explicate the similarities and differences between the two tests, while to put forward another method to explain the scores. The CRTs assessment principles of the quality of Higher Education, which are the theoretical foundation of the next chapter, are presents at the end of the chapter.

The third chapter introduced and distinguished the test scores between NRTs and CRTs by describing and explaining the statistical values, then introduced how to use A, B and Φ values and IRT to guide the test design. It also offered a case analysis as an example.

Chapter Four was an empirical analysis by means of analyzing the TEM4 results about sophomore in Jimei University to further prove that A, B, Φ values and a, b, c values in IRT are better guidance to improve CRLT items designing, then concluded that we must improve the quality of designing test paper in order to identify the different abilities of students, so as to protect the Higher Education Quality Assurance.

The epilogue summarized the innovation of the research and the main conclusions. It also made a simple explanation of some questions needed to continue in study.

Key Words: Modern Educational Measurement Theory; Criterion-reference Language Testing; Test English as Major Grade 4; Application

目 录

第一章 绪 论	1
一、问题的缘起.....	1
二、概念界定.....	3
三、国内外研究综述.....	7
四、研究思路与方法.....	10
第二章 标准参照与标准参照语言测试的理论探讨	13
第一节 标准参照测验的兴起与内涵	13
一、标准参照测验理论的兴起.....	14
二、标准参照测验的概念及内涵.....	15
三、标准参照与常模参照测验的异同.....	18
四、标准参照语言测试的含义.....	23
第二节 标准参照语言测试的编制及使用	25
一、编制的基本原则.....	25
二、项目分析参数.....	25
三、及格的标准水平.....	26
四、效度验证.....	26
五、信度估计.....	27
六、标准参照语言测试的标准化.....	28
第三节 标准参照语言测试与高校标准化考试	29
一、标准化考试.....	29
二、高校标准化考试中的标准参照语言测试.....	31
三、标准参照语言测试的高校教学质量测量法则.....	32
第三章 标准参照语言测试应用理论的分析	34
第一节 经典教育测量理论在标准参照语言测试中的应用	35
一、基本原理.....	35
二、传统项目分析对常模参照语言测试的影响.....	36
三、标准参照语言测试的项目分析.....	38

四、结论.....	46
第二节 现代教育测量理论在标准参照语言测试中的应用原理及优势	47
一、经典教育测量理论的局限性.....	47
二、现代教育测量理论的基本原理.....	49
三、项目反应模型介绍.....	50
四、三种不同模型的优点和不足.....	54
第三节 现代教育测量理论对标准参照语言测试的影响分析	55
----以全国英语学习成绩测验（NEAT）为例	55
一、试题介绍和考后总体分析.....	56
二、听力测验结果分析.....	57
三、结论.....	60
第四章 现代教育测量理论在标准参照语言测试中应用的案例研究	
---以集美大学英语专业四级考试结果为例	62
一、试题总体介绍.....	62
二、具体数据分析.....	63
三、本章小结与启示.....	69
结语.....	71
一、结论与创新.....	72
二、不足.....	73
附录.....	74
参考文献.....	77
后记.....	85

Contents

Chapter One Introduction	1
1. The Origin and Significance of this Research	1
2. Identifying Connotation of Some Concepts.....	3
3. Literature Review.....	7
4. Framework and Method of the Research	10
Chapter Two Basic Principle of Criterion-referenced Tests.....	13
Section 1 The Contents of CRT and its Production Background	13
1. The Historical Background of Producing CRT Theory	14
2. The Implication of CRT	15
3. Identify the Differences between CRT and NRT	18
4. The Connotation of CRLT	23
Section 2 The Basic Thoughts and Methods of CRT	25
1. The Principle of Designing CRT.....	25
2. The Analysis of Item Elements	25
3. Cut-score Criterion.....	26
4. Validity of CRT	26
5. Reliability of CRT	27
6. The Criterion of CRT	28
Section 3 The Standardized CRT and NRT in University	29
1. The Origin and Features of Standardized Tests	29
2. CRLT in Standardized Tests	31
3. The Principles of Higher Education Quality of CRT	32
Chapter Three The Analysis between CRLT and Other Educational Measurements	34
Section 1 Item Analysis in CRT	35
1. Introduction	35

2. The Influence of NRT Item Analysis Based on CEMT	36
3. CRT Item Analysis	38
4. Conclusion	46
Section 2 The Application of Item Response Theory in CRLT	47
1. The Limitation of CEMT	47
2. The Main Contents of IRT	49
3. The Introduction of Three Parameters of IRT	50
4. The Advantages and Limitation of Three Parameters of IRT	54
Section 3 A Case Analysis	55
1. Introduction of National English Achievement Tests	56
2. The Analysis of Listening Test Results	57
3. Conclusion	60
Chapter Four The Application of IRT---A Case Analysis of Jimei	
University	62
1. Introduction	62
2. Analysis of Related-data	63
3. Summary and Conclusion	69
Epilogue	71
Appendix	74
Bibliography	77
Postscript	85

第一章 绪论

高等教育质量是高等教育理论研究中的关键问题，在我国要保证多元化高等教育质量，考试作为衡量教育质量的主要手段，其方式、命题和评分在教学中起着至关重要的作用，科学化的考试能有效推动学生发展，规范和引导教师的教学行为，达到培养规格及教学计划的要求。如何客观、科学地反映和评价每一位学生的专业理论水平和实际应用能力，量化比较不同学校、不同教师、不同教学模式的教学质量和效果，充分发挥考试对改进教学的反馈功能以及考试对教学的指导作用，为社会培养高质量人才，是我国高等教育必须解决的一个重要课题。遗憾的是由于种种历史原因，在废除科举考试以后，中国的考试在理论、技术等方面与西方相比却显得落后了。因此，在高等教育中研究考试理论，开发最合理的考试技术和考试类型，更好地为教育、教学服务，已成为高等教育研究者和考试工作者的历史使命。

一、问题的缘起

中国教育心理界的权威人士，张厚粲教授在她的《考试改革与标准参照测验》的开篇这样说道：“任何教育和心理测验的所得分数都必须有所参照才有实际意义，即测验分数只有与测验以外的某种指标进行比较，才能对被试者的分数给予适当的解释。”^①可见，分数的解释必须有个特定参照，才能全面地进行解释。上世纪80年代初，在我国影响和规模最大的普通高等学校招生全国统一考试中，引进了国外现代教育测量的理论和技术，对传统考试的办法进行了改革，形成了考试科学化、现代化的热潮。高考作为我国高等学校选拔新生的考试，最早它取用的是一种“常模参照测验（Norm-referenced Tests，简称NRT）”，即学生从考试卷上得到的成绩要和所有参加考试的考生成绩相比较，从而反映他在考生总体（常模总体）中的相对位置；随着教育的不断深化，考试改革已不仅局限在考试形式的改革而是向考试内容，考试制度方面改革的深入发展，人们进一步要求完善考试的功能。随着高中新课改的不断深化，任何一位普高学生都需面对高

^①张厚粲，刘昕《考试改革与标准参照测验》[M] 沈阳：辽宁教育出版社，1992，P1

中毕业会考，而这种旨在客观地评价中学生的学业情况，促进学生德、智、体、美、劳的全面发展，并用已定的标准为对照检验学生是否达到培养目标的考试，不是“常模参照性测验”，而是标准参照测验（Criterion-referenced tests，简称CRT）。当参加会考的师生们还用常模参照测验中的理论和技术选定题目，就会发现很好的试题会被弃置不用。因此，靠单一的常模参照测验理论来选定项目（item）、解释分数已远远不能满足当今日益完善的考试制度。

在高校外语教学领域，人才培养也日益需要多元化的评价方式。随着标准参照测验理论的推广，相应的考试改革也在紧锣密鼓地开展。大学英语四、六级（College English Test Band 4&6，简称CET4&6）考试也从原来单一的常模参照语言测试走向标准相关常模参照语言测试（Criterion-related Norm-referenced Language Testing）^①，而英语专业的四、八级（Test English for Major Grade 4&8，简称TEM4&8）考试则已经是标准参照语言测试（Criterion-referenced Language Testing，简称CRLT）。这些变革对于考试理论的研究提出了新的要求。

目前，国内外有较多的书籍介绍常模参照测验的理论和方法，而介绍标准参照测验理论和方法的文章及文献则比较少，介绍CRLT的文献和参考资料则更为稀少；特别是在语言教学越来越重视语言应用目标的今天，在大规模标准化语言考试中，如何把握学习者需要的不同语言应用目标是个难题，如何更好地了解被试者对所学知识的掌握程度（“未掌握、掌握和熟练” non-master, master and mastery）等情况，以及教、考、学的目标如何兼顾都需要作更深入细致地研究，这些不仅需要用到常模参照语言测试理论，更多的需要标准参照语言测试理论，以及其它的测量方法来解决。

高等教育质量评估离不开考试，而考试的评估离不开统计，语言测试的评估也不例外。笔者希望通过本研究来试图分析和区分在高等教育中传统的常模参照测验和标准参照测验的不同，说明高等教育语言教学中，主张以教学大纲为教学指导的标准参照测验（如高校英语专业四、八级考试）的重要性，并进而探讨现代教育测量理论中最具有代表性的项目反应理论（Item response theory，简称IRT）对标准参照语言测试的影响，从而更科学地反映出被试者的能力和检测教师所教的内容，以及学生确实已吸收的适度（fit），以评判被试者的英语能力，

^①杨惠中：大学英语四、六级考试分数解释 [J]《外语界》2001 第 1 期 62~68

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库